Chapter 538

Binary Diagnostic Tests – Clustered Samples

Introduction

A *cluster randomization trial* occurs when whole groups or *clusters* of individuals are treated together. In the two-group case, each cluster is randomized to receive a particular treatment. In the paired case, each group receives both treatments. The unique feature of this design is that each cluster is treated the same. The usual binomial assumptions do not hold for such a design because the individuals within a cluster cannot be assumed to be independent. Examples of such clusters are clinics, hospitals, cities, schools, or neighborhoods.

When the results of a cluster randomization diagnostic trial are binary, the diagnostic accuracy of the tests is commonly summarized using the test *sensitivity* or *specificity*. Sensitivity is the proportion of those that have the condition for which the diagnostic test is positive. Specificity is the proportion of those that do not have the condition for which the diagnostic test is negative.

Often, you want to show that a new test is similar to another test, in which case you use an equivalence test. Or, you may wish to show that a new diagnostic test is not inferior to the existing test, so you use a non-inferiority test.

Specialized techniques have been developed for dealing specifically with the questions that arise from such a study. These techniques are presented in chapters 4 and 5 of the book by Zhou, Obuchowski, and McClish (2002) under the heading Clustered Binary Data. These techniques are referred to as the *ratio estimator* approach in Donner and Klar (2000).

Comparing Sensitivity and Specificity

These results apply for either an independent-group design in which each cluster receives only one diagnostic test or a paired design in which each cluster receives both diagnostic tests. The results for a particular cluster and test combination may be arranged in a 2-by-2 table as follows:

	Diagnostic Test Result		
True Condition	Positive	Negative	Total
Present (True)	T_1	T_0	n_1
Absent (False)	F_1	F_0	n_0
Total	m_1	m_0	N

The hypothesis set of interest when comparing the sensitivities (Se) of two diagnostic tests are

$$H_O: Se_1 - Se_2 = 0 \text{ versus } H_A: Se_1 - Se_2 \neq 0$$

A similar set of hypotheses may be defined for the difference of the specificities (Sp) as

$$H_O: Sp_1 - Sp_2 = 0 \text{ versus } H_A: Sp_1 - Sp_2 \neq 0$$

For each table, the sensitivity is estimated using

$$\hat{S}e = \frac{T_1}{n_1}$$

and the specificity is estimated using

$$\hat{S}p = \frac{F_0}{n_0}$$

The hypothesis of equal difference in sensitivity can be tested using the following *z-test*, which follows the normal distribution approximately, especially when the number of clusters is over twenty.

$$Z_{Se} = \frac{\hat{S}e_1 - \hat{S}e_2}{\sqrt{\hat{V}_{Se1-Se2}}}$$

where

$$\hat{S}e_i = \frac{\sum_{j=1}^{K} n_{1ij}\hat{S}e_{ij}}{\sum_{j=1}^{K} n_{1ij}}$$

$$\hat{V}_{Se_1-Se_2} = \hat{V}ar(\hat{S}e_1) + \hat{V}ar(\hat{S}e_2) - 2\hat{C}ov(\hat{S}e_1, \hat{S}e_2)$$

$$\hat{V}ar(\hat{S}e_i) = \frac{1}{K_i(K_i-1)}\sum_{j=1}^{K_i}\left(\frac{n_{ij}}{\bar{n}_i}\right)^2(\hat{S}e_{ij} - \hat{S}e_i)^2, \quad i = 1, 2$$

$$\hat{C}ov(\hat{S}e_1, \hat{S}e_2) = \frac{1}{K(K-1)}\sum_{j=1}^{K}\left(\frac{n_{1j}}{\bar{n}}\right)^2(\hat{S}e_{1j} - \bar{S}e)(\hat{S}e_{2j} - \bar{S}e)$$

$$\bar{S}e = \frac{\hat{S}e_1 + \hat{S}e_2}{2}$$

$$\bar{n} = \frac{1}{K}\sum_{j=1}^{K} n_{1j}$$

Here we have used K_i to represent the number of clusters receiving test *i.* For an independent design, K_1 may not be equal to K_2 and the covariance term will be zero. For a paired design, $K_1 = K_2 = K$.

Similar results may be obtained for the specificity by substituting *Sp* for *Se* in the above formulae.

Data Structure

This procedure requires four, and usually five, variables. It requires a column containing the cluster identification, the test identification, the result identification, and the actual identification. Usually, you will add a fifth variable containing the count for the cluster, but this is not necessary if you have entered the individual data rather than the cluster data.

Here is an example of an independent-group design with four clusters per test. The Cluster column gives the cluster identification number. The Test column gives the identification number of the diagnostic test. The Result column indicates whether the result was positive (1) or negative (0). The Actual column indicates whether the disease was present (1) or absent (0). The Count column gives the number of subjects in that cluster with the indicated characteristics. Since we are dealing with 2-by-2 tables which have four cells, the data entry for each cluster requires four rows. Note that if a cell count is zero, the corresponding row may be omitted. These data are contained in the BinClust dataset.

BinClust Dataset (Subset)

Cluster	Test	Result	Actual	Count
1	1	0	0	10
1	1	1	0	3
1	1	0	1	2
1	1	1	1	21
2	1	0	0	15
2	1	1	0	2
2	1	0	1	5
2	1	1	1	10
3	1	0	0	23
3	1	1	0	3
3	1	0	1	6
3	1	1	1	31
4	1	0	0	9

Example 1 – Binary Diagnostic Test of a Clustered Sample

This section presents an example of how to analyze the data contained in the BinClust dataset.

Setup

To run this example, complete the following steps:

1 Open the BinClust example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **BinClust** and click **OK**.

2 Specify the Binary Diagnostic Tests – Clustered Samples procedure options

- Find and open the **Binary Diagnostic Tests – Clustered Samples** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Cluster (Group) Variable	**Cluster**
Frequency Variable	**Count**
Diagnostic-Test ID Variable	**Test**
Max Equivalence Difference	**0.2**
Test-Result Variable	**Result**
Test = Positive Value	**1**
Actual-Condition Variable	**Actual**
True = Present Value	**1**
Alpha - Confidence Intervals	**0.05**
Alpha - Hypothesis Tests	**0.05**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary Section

Run Summary Section

Parameter	Value	Parameter	Value
Cluster Variable	Cluster	Rows Scanned	32
Test Variable	Test(1, 2)	Rows Filtered	0
Actual Variable	Actual(+=1)	Rows Missing	0
Result Variable	Result(+=1)	Rows Used	32
Count Variable	Count	Clusters	8

This report records the variables that were used and the number of rows that were processed.

Sensitivity Confidence Intervals Section

Sensitivity Confidence Intervals Section

Statistic	Test	Value	Standard Deviation	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Sensitivity (Se1)	1	0.8214	0.0442	0.7347	0.9081
Sensitivity (Se2)	2	0.7170	0.0518	0.6154	0.8185
Difference (Se1-Se2)		0.1044	0.0681	-0.0291	0.2380
Covariance (Se1 Se2)		0.0000			

Sensitivity: proportion of those that actually have the condition for which the diagnostic test is positive.

This report displays the sensitivity for each test as well as the corresponding confidence interval. It also shows the value and confidence interval for the difference of the sensitivities. Note that for a perfect diagnostic test, the sensitivity would be one. Hence, the larger the values the better.

Specificity Confidence Intervals Section

Specificity Confidence Intervals Section

Statistic	Test	Value	Standard Deviation	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Specificity (Sp1)	1	0.8636	0.0250	0.8147	0.9126
Specificity (Sp2)	2	0.7053	0.0768	0.5547	0.8559
Difference (Sp1-Sp2)		0.1584	0.0808	0.0000	0.3167
Covariance (Sp1 Sp2)		0.0000			

Specificity: proportion of those that actually do not have the condition for which the diagnostic test is negative.

This report displays the specificity for each test as well as corresponding confidence interval. It also shows the value and confidence interval for the difference. Note that for a perfect diagnostic test, the specificity would be one. Hence, the larger the values the better.

Sensitivity & Specificity Hypothesis Test Section

Sensitivity & Specificity Hypothesis Test Section

Hypothesis Test of	Value	Z Value	Prob Level	Reject H0 at 5.0% Level
Se1 = Se2	0.1044	1.5334	0.1252	No
Sp1 = Sp2	0.1584	1.9602	0.0500	Yes

This report displays the results of hypothesis tests comparing the sensitivity and specificity of the two diagnostic tests. The *z* test statistic and associated probability level is used.

Hypothesis Tests of the Equivalence

Hypothesis Tests of Equivalence

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Equivalence at the 5.0% Significance Level
Diff. (Se1-Se2)	0.0803	-0.0076	0.2165	-0.2000	0.2000	No
Diff. (Sp1-Sp2)	0.3032	0.0255	0.2913	-0.2000	0.2000	No

Notes:
Equivalence is concluded when the confidence limits fall completely inside the equivalence bounds.

This report displays the results of the equivalence tests of sensitivity (Se1-Se2) and specificity (Sp1-Sp2), based on the difference. Equivalence is concluded if the confidence limits are inside the equivalence bounds.

Prob Level

The probability level is the smallest value of alpha that would result in rejection of the null hypothesis. It is interpreted as any other significance level. That is, reject the null hypothesis when this value is less than the desired significance level.

Confidence Limits

These are the lower and upper confidence limits calculated using the method you specified. Note that for equivalence tests, these intervals use twice the alpha. Hence, for a 5% equivalence test, the confidence coefficient is 0.90, not 0.95.

Lower and Upper Bounds

These are the equivalence bounds. Values of the difference inside these bounds are defined as being equivalent. Note that this value does not come from the data. Rather, you have to set it. These bounds are crucial to the equivalence test and they should be chosen carefully.

Reject H0 and Conclude Equivalence at the 5% Significance Level

This column gives the result of the equivalence test at the stated level of significance. Note that when you reject H0, you can conclude equivalence. However, when you do not reject H0, you cannot conclude nonequivalence. Instead, you conclude that there was not enough evidence in the study to reject the null hypothesis.

Hypothesis Tests of the Non-inferiority of Test2 Compared to Test1

Tests Showing the Non-inferiority of Test 1 Compared to Test 2

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Non-inferiority at the 5.0% Significance Level
Diff. (Se1-Se2)	0.0000	-0.0076	0.2165	-0.2000	0.2000	Yes
Diff. (Sp1-Sp2)	0.0000	0.0255	0.2913	-0.2000	0.2000	Yes

Notes:
H0: The sensitivity/specificity of Test 1 is inferior to Test 2.
Ha: The sensitivity/specificity of Test 1 is non-inferior to Test 2.
The non-inferiority of Test 1 compared to Test 2 is concluded when the lower c.l. > lower bound.

This report displays the results of noninferiority tests of sensitivity and specificity. The non-inferiority of test 1 as compared to test 2 is concluded if the lower confidence limit is greater than the lower bound. The columns are as defined above for equivalence tests.

Cluster Count Detail Section

Cluster Count Detail Section

Cluster	Test	True Pos (TP)	False Neg (FN)	True Neg (TN)	False Pos (FP)	Total True	Total False	Total Pos	Total Neg	Total
1	1	21	2	10	3	23	13	24	12	36
2	1	10	5	15	2	15	17	12	20	32
3	1	31	6	23	3	37	26	34	29	63
4	1	7	2	9	1	9	10	8	11	19
11	2	25	7	15	6	32	21	31	22	53
12	2	17	3	22	2	20	24	19	25	44
13	2	21	12	16	11	33	27	32	28	60
14	2	13	8	14	9	21	23	22	22	44
Total	1	69	15	57	9	84	66	78	72	150
Total	2	76	30	67	28	106	95	104	97	201

This report displays the counts that were given in the data. Note that each 2-by-2 table is represented on a single line of this table.

Cluster Proportion Detail Section

Cluster Proportion Detail Section

Cluster	Test	Sens. True Pos (TPR)	False Neg (FNR)	Spec. True Neg (TNR)	False Pos (FPR)	Total True	Total False	Total Pos	Total Neg	Prop. Cluster of Total
1	1	0.9130	0.0870	0.7692	0.2308	0.6389	0.3611	0.6667	0.3333	0.1026
2	1	0.6667	0.3333	0.8824	0.1176	0.4688	0.5313	0.3750	0.6250	0.0912
3	1	0.8378	0.1622	0.8846	0.1154	0.5873	0.4127	0.5397	0.4603	0.1795
4	1	0.7778	0.2222	0.9000	0.1000	0.4737	0.5263	0.4211	0.5789	0.0541
11	2	0.7813	0.2188	0.7143	0.2857	0.6038	0.3962	0.5849	0.4151	0.1510
12	2	0.8500	0.1500	0.9167	0.0833	0.4545	0.5455	0.4318	0.5682	0.1254
13	2	0.6364	0.3636	0.5926	0.4074	0.5500	0.4500	0.5333	0.4667	0.1709
14	2	0.6190	0.3810	0.6087	0.3913	0.4773	0.5227	0.5000	0.5000	0.1254
Total	1	0.8214	0.1786	0.8636	0.1364	0.5600	0.4400	0.5200	0.4800	0.4274
Total	2	0.7170	0.2830	0.7053	0.2947	0.5274	0.4726	0.5174	0.4826	0.5726

This report displays the proportions that were found in the data. Note that each 2-by-2 table is represented on a single line of this table.

Example 2 – Paired Design

Zhou (2002) presents a study of 21 subjects to compare the specificities of PET and SPECT for the diagnosis of hyperparathyroidism. Each subject had from 1 to 4 parathyroid glands that were disease-free. Only disease-free glands are needed to estimate the specificity.

The data from this study have been entered in the PET dataset. You will see that we have entered four rows for each subject. The first two rows are for the PET test (Test = 1) and the last two rows are for the SPECT test (Test = 2). Note that we have entered zero counts in several cases when necessary. During the analysis, the rows with zero counts are ignored.

Setup

To run this example, complete the following steps:

1 Open the PET example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **PET** and click **OK**.

2 Specify the Binary Diagnostic Tests – Clustered Samples procedure options

- Find and open the **Binary Diagnostic Tests – Clustered Samples** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Cluster (Group) Variable	**Subject**
Frequency Variable	**Count**
Diagnostic-Test ID Variable	**Test**
Max Equivalence Difference	**0.2**
Test-Result Variable	**Result**
Test = Positive Value	**1**
Actual-Condition Variable	**Actual**
True = Present Value	**1**
Alpha - Confidence Intervals	**0.05**
Alpha - Hypothesis Tests	**0.05**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output

Run Summary Section

Parameter	Value	Parameter	Value
Cluster Variable	Subject	Rows Scanned	84
Test Variable	Test(1, 2)	Rows Filtered	0
Actual Variable	Actual(+=1)	Rows Missing	0
Result Variable	Result(+=1)	Rows Used	53
Count Variable	Count	Clusters	21

Sensitivity Confidence Intervals Section

Statistic	Test	Value	Standard Deviation	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Sensitivity (Se1)	1				
Sensitivity (Se2)	2				
Difference (Se1-Se2)					
Covariance (Se1 Se2)					

Sensitivity: proportion of those that actually have the condition for which the diagnostic test is positive.

Specificity Confidence Intervals Section

Statistic	Test	Value	Standard Deviation	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Specificity (Sp1)	1	0.7843	0.0696	0.6479	0.9207
Specificity (Sp2)	2	0.9020	0.0380	0.8275	0.9764
Difference (Sp1-Sp2)		-0.1176	0.0665	-0.2479	0.0127
Covariance (Sp1 Sp2)		0.0009			

Specificity: proportion of those that actually do not have the condition for which the diagnostic test is negative.

Sensitivity & Specificity Hypothesis Test Section

Hypothesis Test of	Value	Z Value	Prob Level	Reject H0 at 5.0% Level
Se1 = Se2 Sp1 = Sp2	-0.1176	-1.7696	0.0768	No

Hypothesis Tests of Equivalence

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Equivalence at the 5.0% Significance Level
Diff. (Se1-Se2)				-0.2000	0.2000	
Diff. (Sp1-Sp2)	0.1077	-0.2270	-0.0083	-0.2000	0.2000	No

Notes:
Equivalence is concluded when the confidence limits fall completely inside the equivalence bounds.

Tests Showing the Non-inferiority of Test 1 Compared to Test 2

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Non-inferiority at the 5.0% Significance Level
Diff. (Se1-Se2)				-0.2000	0.2000	
Diff. (Sp1-Sp2)	0.1077	-0.2270	-0.0083	-0.2000	0.2000	No

Notes:
H0: The sensitivity/specificity of Test 1 is inferior to Test 2.
Ha: The sensitivity/specificity of Test 1 is non-inferior to Test 2.
The non-inferiority of Test 1 compared to Test 2 is concluded when the lower c.l. > lower bound.

This report gives the analysis of the study comparing PET (Test=1) with SPECT (Test=2). The results show that the two specificities are not significantly different. The equivalence test shows that although the hypothesis of equality could not be rejected, the hypothesis of equivalence could not be concluded either.